2016 15th IEEE International Conference on Machine Learning and Applications

Differentiation and Integration of Machine Learning Feature Vectors

Xinying Mu*, Ana B. Pavel† and Mark Kon‡
*Department of Mathematics and Statistics, Boston University, Boston, MA 02215
Email: xmu@bu.edu
†Graduate Program in Bioinformatics, Boston University, Boston, MA 02215
Section of Computational Biomedicine, Boston University School of Medicine, Boston, MA 02118
Email: anapavel@bu.edu
‡Department of Mathematics and Statistics, Boston University, Boston, MA 02215
Graduate Program in Bioinformatics, Boston University, Boston, MA 02215
Email: mkon@bu.edu

Abstract: This paper presents a new approach to the production of feature maps for the improvement of classification in machine learning. The idea is based on a calculus of differentiation and integration of feature vectors, which can be viewed as functions on a metric space or network. Based on this we propose a novel network-based binary machine learning classifier. We illustrate our method using molecular networks *alone* to distinguish phenotypes, including cancer types and subtypes. We include feature sets derived from disease-specific gene co-expression networks in different cancer data sets using The Cancer Genome Atlas (TCGA) along with other previously published studies. We also illustrate our network-based predictor on another data type, based on infrared spectroscopy of lung cancer tissue.

Keywords: kernel method, classification, gene co-expression networks, cancer

1. Introduction

A major challenge in forming machine learning (ML) classifiers, is the proper choice of feature vectors (FV's) to represent objects. It can be argued that the choice of feature space, feature mapping to a new space, or (equivalently) the choice of a kernel, is as important as the choice of the machine classifier itself [1], [2]. Novel maps producing feature vectors with new characteristics can be a powerful tool in classification. We propose a natural feature map construction that performs calculus-like operations on FV's analogous to the differentiation (and integration) of ordinary functions. For instance feature vectors can be viewed as functions on networks and can be naturally differentiated using the graph Laplacian. Such calculus-like operations on FV's form differentiation feature maps fully transforming feature vectors into new ones. Our goal is to illustrate entirely different feature vectors (i.e., derivatives) that *alone* (i.e., before any combination with original features) can result in classification that equals and sometimes surpasses that of original features.

Feature maps. The usefulness of novel feature maps can be illustrated in the problem of facial recognition ([3], [4]), where the choice of feature map dramatically improves success in an ML task. In the standard representation of FV's in face recognition tasks, as bitmap images with pixel intensities, it has traditionally been difficult to form a machine that directly takes this information and results in appropriate facial recognition. However, the problem of face recognition can be solved quite rapidly with a new feature map. This replaces standard bitmap vector representations of facial images, with feature vectors consisting of distances and ratios of distances among primary facial features. Using a feature vector of 20 to 40 such features can result in very sensitive and even better-than-human facial recognition ([5], [6]).

It is important to develop feature maps that take FV's and canonically map them to other (different) feature vectors that may serve as appropriate starting points for ML algorithms. Forming such feature vectors from the operations of differentiation (and integration) of feature vectors is one such approach.

A simple example. The easiest example of a derivative operation as a feature map occurs when FVs can be identified with functions on the real line. This occurs for example when features have a natural ordinal structure. In the classification of cancer tissue using spectral feature vectors [7], 500 frequencies of infrared light are reflected from an object, leading to a feature vector of 500 numbers ordered from 1 (lowest frequency) to 500 (highest frequency). In such cases, it is already a practice [7], [8] to take derivatives of such feature vectors in the form of first and second difference operations subtracting adjacent features. Our general differentiation methodology reduces in such ordinal cases to naturally produce feature maps similar to the above-mentioned process of differentiation. In other words, such a feature map appears naturally from the above process without any additional algorithmic intervention.

The key observation here is that the FV indices (in this case $1 \dots 500$) have a structure (in this case a simple ordinal one). This structure can also be viewed as a graph, through

978-1-5090-6167-9/16 $31.00 © 2016 IEEE
DOI 10.1109/ICMLA.2016.79

a trivial linear one, in which each node is connected only to one before and one after it. The principle of differentiation in this case can be extended to arbitrary graph structures. In fact the index set of a family of feature vectors can always be given such a graph structure, as discussed below.

The differentiation discussed above can be better understood by noting that a feature vector $x = (x_1, \ldots, x_p)$ can be viewed as a function of its index set $i = 1, \ldots, p$. Such an index set will essentially always have an interesting metric or network structure on which to base a process of differentiation. (figure 1).

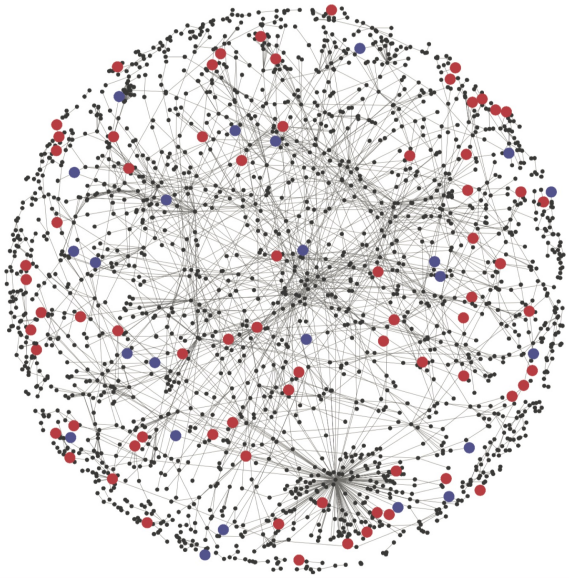

Figure 1. A generic (gene) coexpression network; if feature vector indices are formed by the nodes of the network then the feature vector can be viewed as a function on the network. (This figure was imported from [9], source: http://www.nature.com/msb/journal/v3/n1/full/msb4100138.html)

Our feature map computes the "derivative" of such a function, more specifically its graph Laplacian, or a (positive or negative) fractional power thereof.

In the language of kernels, the choice of a feature map $\Phi(x)$ such as this is equivalent to a re-representation of the kernel $K(x, y)$ being used in the ML task. From that standpoint, the use of our feature map to produce new kernels $K(\Phi(x), \Phi(y))$ from old ones encodes (in computational kernel replacements) the geometric process of differentiating a function.

Network structure. We can illustrate a non-trivial area of application of this idea in computational biology. Let the collection of all genes of a species (e.g. humans) be denoted as V. Note that gene expression array consists of a set of gene-level activations x_i, with $i \in V$ in the set of genes. The feature vector $x = (x_1, \ldots, x_p) = \{x_i\}_{i \in V}$ represents a tissue sample with regard to its individual gene activation levels x_i.

Note also that V, representing the class of genes, can have a network structure based say on a set $E = \{w_{ij}\}$ of (weighted) edges with weights w_{ij}. These weights can represent, say, empirical correlations of gene expression levels obtained from a prior or the present dataset. Alternatively they can represent presence/absence of interaction among the genes i and j at the protein level (e.g. from a database of protein-protein interactions). The resulting graph structure $G = \{V, E\}$ then provides a unique Laplacian matrix operator L that can be applied to individual feature vectors $x = (x_1, \ldots, x_p)$.

This operation $\Phi(x) = Lx$ is the simplest example of this type of feature map. Since the graph Laplacian L is a non-negative self-adjoint operator, we can also take fractional powers L^s of L. Note that L fully generalizes the previously mentioned example of a 'genuine' feature vector derivative, in the case of the spectral feature set $V = \{1, \ldots, 500\}$ with a linear graph structure, where the Laplacian operator forms a true (discrete) second derivative.

Thus the process of feature vector (FV) differentiation occurs when FVs are represented as functions, and ordinary derivative operations (with respect to the underlying graph structure) are performed.

Note in particular that the network structure may depend on the training set of feature vectors (e.g. gene expression vectors) defined on it, but that it may also be entirely independent of the training/testing dataset of the current ML procedure (i.e. it forms prior information). For example in the case of gene expression feature vectors, we can use the protein-protein network structure $G = (V, E)$, where the vertices $v_j \in V$ are genes (or their protein products) and edge weights $w_{ij} \in E$ are $\{0, 1\}$ valued and represent interactions between genes (or proteins). Examples of such prior networks are the abovementioned protein-protein interaction (PPI) networks [10], metabolic pathways [11], transcriptomes [12], and any prior calculated gene co-expression networks [13].

Standard vs. differentiated feature vectors. Consider now the case of networks trained based on correlations of features in the training set D. The new set of features obtained from the derivative feature map $\Phi(x) = Lx$ can be viewed as encompassing a different type of information than the original feature vector x. Indeed, the original family of feature vectors in training and testing presents information regarding the *locations* of (sample) feature vectors typically belonging to a class A to be identified, allowing a *geometrical* separation of two classes A and B in the original feature space $\mathcal{F}$.

The feature vector Lx however encodes information about the *relative* locations of individual features to each other. Indeed, in the above simple example of a linear ordered network of 500 spectral features, the Laplacian of a feature vector x represents a *second difference* of the intensity of feature x_i relative to its network neighbors (in this case $x_{i\pm1}$). This can indicate if the feature vector x fits into a pattern of network variation encoded in the training set say for class A, rather than that for class B. This is a qualitatively different type of information, though it is directly derived from x via the feature map $\Phi : x \to Lx$. So when a feature vector is classified, relationships among

its components are being used to determine its class. Differentiation of feature vectors allows the classification of test vectors x according to whether they fit into the inferred network of relationships of features of a given class A, obtained from the training set D.

Application to molecular network-based classification. This general method can be applied to a range of real-life applications. Computational biology is currently an expanding area that benefits from the explosion of molecular data types being studied. Conventional classifiers identify gene expression patterns by treating genes as independent features. However this may not always capture real biological processes - biological systems are complex with all molecular entities dynamically interacting in the cell. In some applications below we seek to use the network of gene interactions and correlations as a source of information using network information to stratify cancer types and subtypes.

The edges of a gene network reveal new features distinct from direct expression values, giving a new level of information that can be used for classification of test samples. Currently gene expression signatures are most commonly exploited for classifiers using different statistical and ML techniques, such as logistic regression, random forests, support vector machines, etc. Gene expression classifiers have been developed for lung cancer [14], breast cancer [15], colo-rectal cancer [16] and other cancer types. However, most times these classifiers tend to overfit due to noise in the data that generates false patterns [17].

Networks and groupings of genes have nevertheless become important in various types of classification tasks. Hofree [18] proposes a network based approach for stratification of tumor mutations. The authors test their method in ovarian, uterine and lung cancer cohorts from TCGA. For each tissue, they identify subtypes that are predictive of clinical outcomes such as patient survival, response to therapy or tumor histology.

Previous work has shown that gene expression classifiers can be improved by exploiting gene interaction information. The authors in [19] show that regularizing (smoothing) micro-array expression values defined on gene sets with known prior network or metric structures improve prediction.

Below we test the predictor alone and evaluate its performance in multiple cancer data sets, including three TCGA cancer types (breast cancer [20], lung adenocarcinoma [21] and lung squamous cell carcinoma [22]), a collection of benchmark cancer data sets [23] and infrared spectroscopy lung cancer data [8].

2. Methods

Overview. The idea of our approach to predictive classification in machine learning can be illustrated in a simple way. In the case of two classes A and B to be distinguished, assume that there is a training set $D = D_1 \cup D_2$, with $D_1 = \{x^k\}_{k=1}^{n_1}$ consisting of training data in class 1, and $D_2 = \{x^k\}_{k=n_1+1}^{n}$ consisting of training data in class 2.

The classification method starts by learning the empirical correlation matrix $\Sigma^1 = (\sigma_{ij}^1)_{i,j}$, with σ_{ij}^1 the correlation of features x_i and x_j in dataset D_1. Similarly, matrix Σ^2 identifies the correlation structure of the same set of features, now in datset D_2. To clarify notation, the sample data point x^k represents the gene expression vector of sample (e.g. patient) k in dataset D, with component x_i^k representing the expression level of gene i in sample k.

The above empirical network structures based on the correlation networks Σ^1 and Σ^2 of the datasets D_1, D_2 will be used in a simple way to score test vectors $x = (x_1, \ldots, x_p)$ with regard binary classification as to membership in two classes, here denoted as 1 and 2.

As illustrated in figure 2, gene expression correlation networks provide condition-specific patterns that can be used to distinguish classes. In the figure these consist of correlation patters for gene expression in a group of cancer patients and a group of control patients, for the same gene set.

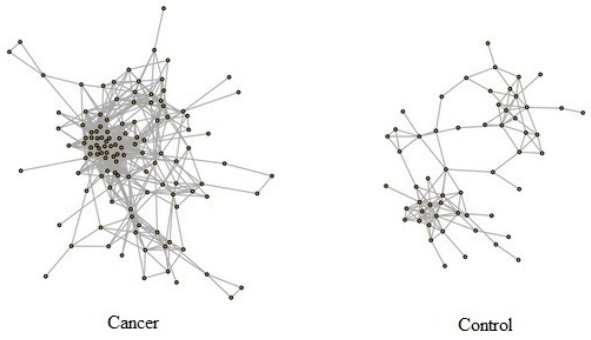

Figure 2. For this cancer/normal dataset, the correlation network built from a disease group differs from that built from a control group for the same 100 genes. These networks were generated using the Diffuse Large B-Cell Lymphoma (DLBCL) data from [23].

Formalization of the model. We will view a feature vector $x = (x_i)_{i=1}^p$ as a function (a *feature function*) $f_x(q)$ on the index set, $q \in 1, \ldots, p$. Thus $f_x(q)$ represents a function on the network $G = (V, E)$ with V the set of genes (or equivalently indices $1, \ldots, p$) and $E = \{w_{ij}\}_{i,j \in V}$ consisting of edge weights between the vertices in V.

We in fact have two network structures on the set V of genes, one trained from the dataset D_1 (class 1) and one trained from dataset D_2 (class 2), and denote the network structures by $G_1 = (V, E_1)$ and $G_2 = (V, E_2)$. Viewed as notions of distance or proximity, these two networks effectively define two different geometries on the space V of indices, which will be our viewpoint below.

Note we expect feature functions $f(q)$ on V (arising from feature vectors x, with $f_x(q) = x_q$), to be adapted to different geometries on V depending on their classes. Thus for example we expect a feature function $f_x(q)$ from a sample x in class 1, to be adapted to the geometry (network structure) G_1 on V, in a sense clarified below.

More specifically, let q and r be 'nearby' points in V, according to the 'geometry' G_1. Then we would expect

$f_x(q)$ and $f_x(r)$ to be close to each other if x represents a feature vector from class 1, and further from each other if it is from class 2.

This closeness condition defines the notion that f_x is adapted to the geometry G_1 on V. Recalling the notion of 'smoothness' of a function f_x on a space V in geometric terms, the condition intuitively states that f_x is 'smooth' on V with geometry G_1. More specifically we expect a feature function f_x defined on V that is from class 1 to be 'smooth' on V when if it is equipped with geometry G_1, more so than when f_x is from class 2 (in which case f_x will be more adapted to geometry G_2).

Practically this will mean that the G_1-graph Laplacian (derivative) $L_1 f_x$ for feature vectors x in class 1 will be smaller (in the square integral sense) than feature vectors x from class 2. Recall that for the graph G_1 (with $|G_1| = p$ vertices), its $p \times p$ Laplacian matrix is defined as

$$L = D - W$$

Here the diagonal matrix D has entries $D_{ii} = \sum_{j \in V} w_{ij}$ representing the sum of weights connected to vertex i, while W is the connectivity matrix, with $W_{ij} = w_{ij}$.

Thus we train two networks based on the two classes to be separated, and use the networks to differentiate test feature vectors between the two classes. In the language of feature vectors (rather than feature functions), let x and y be feature vectors from classes 1 and 2. Assume we differentiate (take the Laplacians of) these vectors under the different network structures induced by classes 1 and 2. Since x is from class 1, it is naturally adapted to the metric structure of class 1. Thus the square norm of the class 1 network Laplacian

$$||L_1 x||^2 = \sum_i (L_1 x)_i^2$$

(with respect to the metric structure of network 1) will be smaller than that with respect to the metric structure of class 2. Conversely, the Laplacian norm $||L_2 y||^2$ of y with respect to the class 2 network structure will be smaller than that with respect to the class 1 network structure.

This measure of adaptation of feature vectors to the relative metric (network) structures of their feature spaces is a tool forming new feature vectors to separate objects in the two classes. In effect this forms a feature map that yields new and differently structured feature vectors Lx from old ones x, for improved machine learning recognition, in particular when used to augment standard features.

The computational implementation (described below) is very simple in principle. Derivatives of test feature vectors from an unknown class can be immediately taken by applying the two graph Laplacians $L_i \ \ (i = 1, 2)$ relative to the network structures trained from the two classes. In training, a threshold is formed effectively separating the mean values of the two Laplacians $\langle x, L_1 x \rangle - \langle x, L_2 x \rangle$ among training feature vectors x in the two classes (see next section for details). Here $\langle x, y \rangle$ represents the ordinary (multiply and sum) dot product of x and y. In testing, vectors which lie above the threshold are classified as being in class 2, while those below are classified as 1

Implementation: In more detail, given a trainining set $D = D_1 \cup D_1$ consisting of data D_1 in class 1 and D_2 in class 2, we first build gene interaction networks across the training samples of each of the two classes. To capture the network of gene interactions, we compute the Spearman rank correlation coefficient σ_{ij} between expression measurements of each pair of genes i, j. This estimates how gene pairs influence each others' expression levels.

For each feature vector $f_x = f = (f_1, \cdots, f_p)$, we define the smoothness (with respect to the network geometry) by:

$$y(f) = \sum_{i,j} (f_i - f_j)^2 \cdot w_{ij} = f^T \cdot L \cdot f = \langle f, Lf \rangle \quad (1)$$

where L is the Laplacian matrix, and w_{ij} is the connection weight between feature node i and j. Note that here the feature function $f = f_x$ and feature vector x are used interchangeably. In addition w_{ij} is the correlation coefficient between features i and j in the given class (1 or 2). The second identity in the above equation (expressing $y(f)$ in terms of the Laplacian) is easy to verify.

More generally, the Laplacian can be replaced by its fractional powers,

$$L^* = (L + \lambda)^s \quad (2)$$

with λ, s parameters. The small parameter λ plays the role of a regularization parameter for negative powers s.

Then we define the class separation threshold as the average of the two relative smoothness values, computed as the difference between smoothnesses under the two networks. The classification threshold γ can be defined as

$$\gamma = \frac{n_1 \cdot [\bar{y}_1(f_1) - \bar{y}_2(f_1)] + n_2 \cdot [\bar{y}_1(f_2) - \bar{y}_2(f_2)]}{n_1 + n_2}, \quad (3)$$

where n_1 and n_2 are numbers of samples in class 1 and class 2 respectively, and e.g., $\bar{y}_1(f_2)$ represents the mean value of smoothness for feature vectors in class 2 on network G_1, so that $\bar{y}_1(f_2) = \sum_{f \in \text{class2}} y_1(f)/n_2$.

The method is somewhat sensitive to class size, since it needs similar numbers of samples to build comparable networks for the two classes. The above threshold definition, adjusted for sample size, can reduce the false predictions from unbalanced datasets.

The algorithm consists of four steps:
Step 1. Generate the correlation network G_k for each class k.
Step 2. Compute the smoothness for each sample according the network for each class.
Step 3. Find the separation threshold of smoothness values for the two classes, as defined in (3).
Step 4. Test prediction accuracy by comparing smoothness differences between the two networks, under the threshold γ.

3. Applications

3.1. Performance on benchmark classification algorithms

We first tested our approach on gene expression data from four of the benchmark cancer data sets studied in [23]. We emphasize that (table 1), the purpose of the present results for the network method is to benchmark classification based on the single feature γ *alone*, in order to gauge the strength of network geometry as a classifier. Here γ is as in (3) above, representing the difference in smoothness of the feature vector relative to two network geometries. The feature γ can of course be augmented and combined with additional features, for example the original feature vector f, and other functions of its Laplacian $L \cdot f$, in more general approaches.

Data in all tables represent accuracy levels in percent.

TABLE 1. LOOCV ACCURACY OF CLASSIFIERS FOR BINARY CLASS EXPRESSION DATASETS

algorithms	Network	TSP	k-TSP	SVM	k-NN
DLBCL	82.73	98.1	97.4	97.4	84.42
Colon	70.97	91.1	90.3	82.26	74.19
Prostate 3	96.97	97	97	100	87.88
Lung	93.92	98.3	98.9	99.45	98.34
Average	86.15	88.26	92.01	91.18	81.63

3.2. TCGA breast and lung cancer data

In this section we test the proposed network-based classifier on three gene expression data sets from The Cancer Genome Atlas (TCGA): breast invasive carcinoma (BRCA) [20], lung squamous cell carcinoma (LUSC) [22] and lung adenocarcinoma (LUAD) [21]. Gene expression levels estimate the abundance of RNA transcripts that will ultimately be translated into proteins. We use gene expression profiled by RNA-Sequencing on the Illumina HiSeq platform. The RPKM values were normalized as $log_2(RPKM + 1)$.

We compared the results of the network-based classifier with two other classifiers that use gene expression values as independent classification features (SVM and K-NN). To evaluate the performance, we computed the mean accuracy of 10-fold cross validation for each predictor (table 2).

3.2.1. Breast cancer. We used RNA-sequencing gene expression data for 113 cancer and 113 normal patients with breast invasive carcinoma (BRCA) [20]. The BRCA RNA-sequencing data consists of 1107 cancer and 113 normal patients; it provides the largest set of normal samples with gene expression measurements in TCGA. For this dataset, we used all the normal samples available and a subset of the same size of the cancer samples.

3.2.2. Lung cancer. Here we considered RNA-sequencing gene expression data for 51 normal and 501 cancer patients with lung squamous cell carcinoma (LUSC) [22], and for 59 normal and 528 cancer patients with lung adenocarcinoma (LUAD) [21].

As expected, the normal and cancer classes are well differentiated for all three cancer types. All three predictors perform similarly well with a higher than 90% accuracy (table 2).

TABLE 2. CROSS VALIDATION ACCURACY OF CLASSIFIERS FOR BINARY CLASS TCGA DATA SETS

Data	Network	SVM	K-NN
BRCA vs. Normal	97.37	100	91.09
LUAD vs. Normal	98.23	99.07	96.52
LUSC vs. Normal	98.31	99.55	99.45

3.3. Infrared spectroscopy data on lung cancer tissue

Spectral histopathology (SHP) works on the principle that all biochemical components have distinct fingerprints in the form of infrared spectral signatures observable via infrared spectroscopy [24], [7]. When observed through an infrared microscope, objects smaller than a human cell can be identified and their spectra can be acquired.

The training portion of this dataset contains 2,000 pixels in squamous cell carcinoma (SqCC) and 2,000 in adenocarcinoma (ADC) from 182 patients, with 501 wavenumber frequency features. The test data have the same form, from 49 patients [7].

As above, we compare the performance of the network-based predictor with the results of SVM and K-NN machines (table 3). The accuracy results in the table are averaged under 10-fold cross validation. All three predictors perform well; however in this case the network-based approach alone outperforms SVM and K-NN.

TABLE 3. CROSS-VALIDATION ACCURACY OF BINARY CLASSIFIERS FORH INFRARED SPECTROSCOPY LUNG CANCER DATA

Data	Network	SVM	K-NN
SqCC vs. ADC	89.3	87.5	78.1

4. Conclusions

In this paper we have demonstrated the usefulness of differentiation of (taking derivatives of) feature vectors in ML problems, with respect to network structures from within their training data or from prior outside data. The process is implemented by computing the graph Laplacian L of feature vectors, and more generally the fractional Laplacian $(L + \lambda)^s$, where s can be fractional. Future work will incude choices of negative values for s, which will imply

an *integration* operation, that will smooth feature vectors, possibly eliminating noise. We have shown that a single feature γ representing smoothness with respect to network structure can be used to separate classes of feature vectors successfully. This new feature type has been demonstrated successfully in some binary classification problems involving gene expression and other molecular-level predictors. Our model captures the information of gene regulation in the form of dynamical variations caused by gene interactions. This new level of information describes real biological processes that may be complementary to gene expression measurements and as independent features.

Our approach may serve as a valuable adjunct for biomarker discovery and identification of disease specific molecular interactions.

Acknowledgments

The authors would like to acknowledge the TCGA Research Network and Cireca Theranostics. The results shown here are in part based upon data generated by the TCGA Research Network: http://cancergenome.nih.gov/. We also acknowledge The Graduate Program in Bioinformatics at Boston University.

References

[1] B. Schölkopf and A. J. Smola, *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2002.

[2] N. Cristianini and J. Shawe-Taylor, *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000.

[3] X. Jia and M. S. Nixon, "Extending the feature vector for automatic face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 12, pp. 1167–1176, Dec 1995.

[4] C. Geng and X. Jiang, "Face recognition using sift features," in *Proceedings of the 16th IEEE International Conference on Image Processing*, ser. ICIP'09. Piscataway, NJ, USA: IEEE Press, 2009, pp. 3277–3280. [Online]. Available: http://dl.acm.org/citation.cfm?id=1819298.1819645

[5] A. Mian, Y. Hu, R. Hartley, and R. Owens, "Image set based face recognition using self-regularized non-negative coding and adaptive distance metric learning," *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 5252–5262, Dec 2013.

[6] J. Križaj, V. Štruc, and N. Pavešic, "Adaptation of sift features for face recognition under varying illumination," in *MIPRO, 2010 Proceedings of the 33rd International Convention*, May 2010, pp. 691–694.

[7] X. Mu, M. Kon, A. Ergin, S. Remiszewski, A. Akalin, C. M. Thompson, and M. Diem, "Statistical analysis of a lung cancer spectral histopathology (shp) data set," *Analyst*, vol. 140, no. 7, pp. 2449–2464, 2015.

[8] A. Akalin, X. Mu, M. A. Kon, A. Ergin, S. H. Remiszewski, C. M. Thompson, D. J. Raz, and M. Diem, "Classification of malignant and benign tumors of the lung by infrared spectral histopathology (shp)," *Lab Invest*, vol. 95, no. 4, pp. 406–421, Apr 2015, pathobiology in Focus. [Online]. Available: http://dx.doi.org/10.1038/labinvest.2015.1

[9] X. Lu, V. V. Jain, P. W. Finn, and D. L. Perkins, "Hubs in biological interaction networks exhibit low changes in expression in experimental asthma," *Molecular Systems Biology*, vol. 3, no. 1, 2007. [Online]. Available: http://msb.embopress.org/content/3/1/98

[10] L. Matthews, G. Gopinath, M. Gillespie, M. Caudy, D. Croft, B. de Bono, P. Garapati, J. Hemish, H. Hermjakob, B. Jassal, A. Kanapin, S. Lewis, S. Mahajan, B. May, E. Schmidt, I. Vastrik, G. Wu, E. Birney, L. Stein, and P. D'Eustachio, "Reactome knowledgebase of human biological pathways and processes," *Nucleic Acids Research*, vol. 37, no. suppl 1, pp. D619–D622, 2009.

[11] M. Kanehisa and S. Goto, "Kegg: Kyoto encyclopedia of genes and genomes," *Nucleic Acids Res*, vol. 28, no. 1, pp. 27–30, Jan 2000, gkd027[PII]. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC102409/

[12] V. Matys, E. Fricke, R. Geffers, E. Gößling, M. Haubrock, R. Hehl, K. Hornischer, D. Karas, A. E. Kel, O. V. Kel-Margoulis, D.-U. Kloos, S. Land, B. Lewicki-Potapov, H. Michael, R. Münch, I. Reuter, S. Rotert, H. Saxel, M. Scheer, S. Thiele, and E. Wingender, "Transfac: transcriptional regulation, from patterns to profiles," *Nucleic Acids Research*, vol. 31, no. 1, pp. 374–378, 2003.

[13] B. Zhang, S. Horvath, B. Zhang, and S. Horvath, "A general framework for weighted gene coexpression network analysis," in *Statistical Applications in Genetics and Molecular Biology 4: Article 17*, 2005.

[14] D. H. Whitney, M. R. Elashoff, K. Porta-Smith, A. C. Gower, A. Vachani, J. S. Ferguson, G. A. Silvestri, J. S. Brody, M. E. Lenburg, and A. Spira, "Derivation of a bronchial genomic classifier for lung cancer in a prospective study of patients undergoing diagnostic bronchoscopy," *BMC Medical Genomics*, vol. 8, no. 1, pp. 1–10, 2015. [Online]. Available: http://dx.doi.org/10.1186/s12920-015-0091-3

[15] L. J. van 't Veer, H. Dai, M. J. van de Vijver, Y. D. He, A. A. M. Hart, M. Mao, H. L. Peterse, K. van der Kooy, M. J. Marton, A. T. Witteveen, G. J. Schreiber, R. M. Kerkhoven, C. Roberts, P. S. Linsley, R. Bernards, and S. H. Friend, "Gene expression profiling predicts clinical outcome of breast cancer," *Nature*, vol. 415, no. 6871, pp. 530–536, Jan 2002. [Online]. Available: http://dx.doi.org/10.1038/415530a

[16] F. Bertucci, S. Salas, S. Eysteries, V. Nasser, P. Finetti, C. Ginestier, E. Charafe-Jauffret, B. Loriod, L. Bachelart, J. Montfort, G. Victorero, F. Viret, V. Ollendorff, V. Fert, M. Giovaninni, J.-R. Delpero, C. Nguyen, P. Viens, G. Monges, D. Birnbaum, and R. Houlgatte, "Gene expression profiling of colon cancer by dna microarrays and correlation with histoclinical parameters," *Oncogene*, vol. 23, no. 7, pp. 1377–1391, 0000. [Online]. Available: http://dx.doi.org/10.1038/sj.onc.1207262

[17] A. V. Tinker, A. Boussioutas, and D. D. L. Bowtell, "The challenges of gene expression microarrays for the study of human cancer," *Cancer Cell*, vol. 9, no. 5, pp. 333–339, 2016. [Online]. Available: http://dx.doi.org/10.1016/j.ccr.2006.05.001

[18] M. Hofree, J. P. Shen, H. Carter, A. Gross, and T. Ideker, "Network-based stratification of tumor mutations," *Nat Meth*, vol. 10, no. 11, pp. 1108–1115, Nov 2013. [Online]. Available: http://dx.doi.org/10.1038/nmeth.2651

[19] Y. Fan, M. Kon, S. Kim, and C. DeLisi, "Smoothing gene expression using biological networks," in *Machine Learning and Applications (ICMLA), 2010 Ninth International Conference on*, Dec 2010, pp. 540–545.

[20] The Cancer Genome Atlas Research Network and others, "Comprehensive molecular portraits of human breast tumours," *Nature*, vol. 490, no. 7418, pp. 61–70, 2012.

[21] ——, "Comprehensive molecular profiling of lung adenocarcinoma," *Nature*, vol. 511, no. 7511, pp. 543–550, 2014.

[22] ——, "Comprehensive genomic characterization of squamous cell lung cancers," *Nature*, vol. 489, no. 7417, pp. 519–525, 2012.

[23] A. C. Tan, D. Q. Naiman, and D. G. Lei Xu, Raimond L. Winslow, "Simple decision rules for classifying human cancers from gene expression profiles," *Bioinformatics*, vol. 21, no. 20, p. 3896–3904, 2005.

[24] M. Diem, A. Mazur, K. Lenau, J. Schubert, B. Bird, M. Miljković, C. Krafft, and J. Popp, "Molecular pathology via ir and raman spectral imaging," *Journal of biophotonics*, vol. 6, no. 11-12, pp. 855–886, 2013.